

Published in final edited form as:

Eng Appl Artif Intell. 2007 December ; 20(8): 1070–1085.

Genetic Programming of Conventional Features to Detect Seizure Precursors

Otis Smart *

gte851r@mail.gatech.edu, Ph.D. Candidate, Georgia Institute of Technology, Intelligent Control Systems Laboratory, Biomedical Engineering Research Group, 813 Ferst Drive, N.W., Atlanta, GA 30332

Hiram Firpi, and

hfirpi@compbio.iupui.edu, Post-doctoral Fellow, Indiana University School of Medicine, Center for Computational Biology and Bioinformatics, 714 North Senate Avenue, Suite 250, Indianapolis, IN 46202-5122

George Vachtsevanos

gfv@ece.gatech.edu, Professor Georgia, Institute of Technology School of Electrical and Computer Engineering, Van Leer Building, 777 Atlantic Drive, Atlanta, GA 30332

Abstract

This paper presents an application of genetic programming (GP) to optimally select and fuse conventional features (C-features) for the detection of epileptic waveforms within intracranial electroencephalogram (IEEG) recordings that precede seizures, known as seizure-precursors. Evidence suggests that seizure-precursors may localize regions important to seizure generation on the IEEG and epilepsy treatment. However, current methods to detect epileptic precursors lack a sound approach to automatically select and combine C-features that best distinguish epileptic events from background, relying on visual review predominantly. This work suggests GP as an optimal alternative to create a single feature after evaluating the performance of a binary detector that uses: 1) genetically programmed features; 2) features selected via GP; 3) forward sequentially selected features; and 4) visually selected features. Results demonstrate that a detector with a genetically programmed feature outperforms the other three approaches, achieving over 78.5% positive predictive value, 83.5% sensitivity, and 93% specificity at the 95% level of confidence.

Keywords

genetic programming; C-features; feature-selection; feature-fusion; binary detection; epilepsy; epileptic precursors; IEEG

1. Introduction

Epilepsy afflicts over 60 million people worldwide and more than 25% of those persons experience seizures that violently emerge throughout the entire brain and cannot be controlled by conventional therapies, such as surgery, because the disease manifests in multiple locations or is too complex to localize by doctors. Although new therapeutic options, such as electrical

*Corresponding Author Phone: 404.894.9356 Fax: 404.894.4130.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

stimulation, appear to be a viable alternative for treating patients with intractable seizures, a reliable means to automatically isolate the most considerably diseased regions of brain is still a challenge that needs to be met in order to guide innovative therapies.

It has been suggested that epileptic seizure precursors—abnormal, seizure-like waveforms preceding a seizure—within electrographic recordings of brain activity known as an electroencephalogram (EEG) may localize regions important to seizure generation. Despite the large number of attempts over decades to accurately detect precursors, previous schemes to detect precursors have exhibited inconsistent performance (i.e., 74–94% sensitivity, 84–90% specificity, 54–88% selectivity) across small databases of testing records (Clavango 2000, Hassanpour 2003, Hassanpour 2004, Ossadtchi 2002, Pon 2002, Adjouadi, 2004, Jones, 1996, Ko, 1998, Tarassenko, 1998, Van Hoey, 1998, Van Hese, 2003, Wilson, 1991, Zhang, 1999) without reporting level of confidence in those results. For instance, initial work in precursor-detection was focused on detecting spikes and sharp-wave transients within the EEG, yet the various approaches for detection and features for extraction have not cultivated connections between the precursors and localized epileptogenic zones (Clavango 2000, Hassanpour 2003, Hassanpour 2004, Ossadtchi 2002, Pon 2002, Adjouadi, 2004, Jones, 1996, Ko, 1998, Tarassenko, 1998, Van Hoey, 1998, Van Hese, 2003, Wilson, 1991, Zhang, 1999). On the other hand, the recent efforts of a few groups have reported more promising findings concerning precursors other than spikes and sharp-waves. Staba et al. (2002) detected and analyzed interictal oscillations between 80–500 Hz within the hippocampus and entorhinal cortex of humans with focal epilepsy. Though only reporting a sensitivity of 84% for their detector and arbitrarily choosing a feature correlated to a measure of energy, their work demonstrated an association between the oscillations and epileptogenic regions. Niederhauser et al. (2003) detected synchronous bursts of activity between 20–40 Hz in humans with focal epilepsy in the medial temporal lobe. While reliable detection was achieved in five patients with unilateral seizure onsets, the method did not recognize the precursor in five other patients more generalized onsets. In 2004 and 2005, Smart et al. investigated epileptiform oscillations between 60–100 Hz before generalized seizure onsets in humans, reporting that the oscillations localized regions and moments in time before the earliest electrographic change in depth recordings in 77% and 61% of the observed onsets, respectively; however, the recommended oscillation-detector was only 68% selective.

Overall the “state-of-the-art” in precursor detection lacks robust choices for the selection and combination of quality features that separate precursors from non-precursors. This work resolves this persistent issue by adopting an incredibly versatile and effective tool for feature-selection and feature-fusion: genetic programming (GP). Surprisingly no work to date uses GP to facilitate the detection of seizure precursors, although genetic programming has been applied to the detection of epileptic seizures with considerable success (Firpi, 2005; Firpi, et al., 2005a; Firpi, et al., 2005b). This paper proposes an automated methodology to detect seizure precursors, relying on optimal feature extraction via a genetically programmed feature (GP-feature). In Section 2, a brief explanation of the theory of GP is provided. In Section 3, the suggested methodology for robust precursor-detection is described in detail. In Section 4, the proposed method is implemented using popular approaches for feature-selection (without feature-fusion), and an analysis of the best technique is performed. Finally, Section 5 contains concluding remarks for the presented and future work.

2. Theory of Genetic Programming (GP) Algorithm

Genetic programming, which was established by Koza (Koza, 1989, 1992, 1994) is a general-purpose, global search optimization procedure that computes a solution by imitating three processes of natural evolution described by Darwin: *selection*, *crossover*, *mutation*, and *survival*. More specifically, the algorithm initializes a set (*population*) of solutions

(*individuals*) of size P with each element representing a mathematical operation on the input of the GP in the form of a tree structure; computes per individual a measure of fitness to an ideal solution; and executes the selection, crossover, mutation, and survival stages to create new populations while either maximizing or minimizing the *fitness* of the best individual, depending on the formulation of the problem.

Although many variations of the selection, crossover, mutation, and survival operations in GP exist, the basic concept of each process is constant. In the selection stage, a subset of the current population (*intermediate population*) is chosen based upon the fitness of the individual; in the crossover stage, the GP creates new individuals (*children*) using combinations of a pair of individuals (*parents*) from the intermediate population that will belong to a new population; the mutation stage introduces diversity into the new population by randomly altering the makeup of a subset of individuals in the new population; and the survival stage simply selects the fittest individuals from the new population, creating a new initial population of size P for subsequent iterations of the GP. The algorithm terminates upon reaching a predefined number of *generations* (iterations) or a level of fitness among the present population.

The following figures elucidate the operation and application of a GP. Figure 1 depicts the overall genetic programming procedure, and Figure 2 is an example of a result from a GP program for a particular problem. Figures 3 and 4 demonstrate the crossover and mutation operations, respectively. According to an appropriate fitness measure, the GP algorithm heuristically finds an optimal solution, or individual, from a set of potential solutions. The individual is a tree structure that contains two types of nodes: *functions* that receive a number of arguments on which they operate and *terminals*, which are values selected from the input.

For instance, if a three-dimensional regression problem were posed in the framework of GP, the tree in Figure 2 would be an estimate of the “best” regression equation, likely in a least-square error sense, and the terminals of this tree would represent known abscissae of the estimated curve. If instead the problem were *feature-fusion*, a tree would correspond to a mapping from the original feature-space spanned by the selected features into the feature-space of the composite feature, and the terminals would be the selected features. Appendix A provides more detail on applying GP to *feature-fusion*.

3. Methodology

3.1 Data Collection

Intracranial electroencephalogram (IEEG) recordings from two patients—one adult and one pediatric—with neocortical epilepsy who underwent evaluation with EEG electrodes for resective surgery were studied. The data for each patient were collected using a digital, 64-channel, 12-bit, Nicolet BMS-5000 (Nicolet Biomedical Inc., Madison, WI) epilepsy monitoring system. The epilepsy monitoring systems are housed at The Mayo Clinic in Rochester, MN and the Children’s Hospital of Philadelphia. Referentially recorded EEG was band-passed filtered from 0.1–100 Hz and digitized at 200 Hz before archiving to CD-ROM for later processing. Bipolar electrode montages were used to reduce common mode artifact, and a digital 60 Hz notch filter was implemented to eliminate line noise.

Seizure onsets for the entire data archive were marked by two board-certified epileptologists. The doctors determined the onset times for each seizure by looking backwards in the record for the earliest EEG change (EEC) from the baseline activity and identifying a clear electrographic seizure discharge at the time of “unequivocal EEG onset (UEO).” The seizure onset zone (SOZ) was chosen as the electrode location(s) with the earliest seizure onset and represented the most notably diseased sites of epileptic brain. On the other hand, onsets for seizure precursors—in this case high-frequency epileptiform oscillations (HFEO’s)—were

marked in only a few (40) records with seizures in the SOZ for a three-minute interval in each patient. Marking oscillations across sixty-four channels of available recordings in several patients would have been an impractical task, especially since the duration of the oscillations was on the order of milliseconds. Therefore, the epileptologists marked and scored with a rating from 1 (low quality event) to 10 (high quality event) only a small collection of onset times for HFEO's for preliminary analyses.

3.2 Data Selection

From the above collection of data, three-minute clippings of continuous IEEG from the SOZ with epileptic oscillations that manifested intermittently within the electrographic "background" were made. Using these clippings, two sets of data were prepared for the suggested approach to detect HFEO's. A set of *training* data was used to create a feature with high separation of classes for the feature extraction module via a genetic programming (GP) algorithm, generate target vectors for a classifier based on values of that feature, and cross-validate the classifier. A set of *testing* data was used to verify that the method could work well given data that is not known a priori, guide any potential improvements on the method, and found early observations about high-frequency epileptiform oscillations.

3.3 Signal Processing

All work toward designing an automated HFEO-detector was done using MATLAB version 7.0.4.365 (R14) software on a Dell Dimension 8300 desktop equipped with an Intel Pentium 4 processor (3.2 GHz) and 1 GB of RAM. The four major modules for the detector included a pre-processing stage to accentuate potential HFEO's, a processing stage to extract and transform a chosen set of classical features, a processing stage to classify feature values as either an HFEO or a non-HFEO, and a post-processing stage to reduce the detection of spurious events.

3.4 Pre-processing

As remarked by Worrell et al. (2004), the frequency content of the high-frequency epileptiform oscillations could range from 60 to 100 Hz in adult patients with neocortical epilepsy. In addition, Marsh (2005) has noticed that similar epileptic waveforms in pediatric patients oscillate within 50 to 85 Hz. Thus band-pass filtering the IEEG to improve the signal-to-noise ratio between the activity of an HFEO and non-HFEO, or a signal equivalent to typical electrographic background, was employed. For the band-pass filtering, a Chebychev filter with a small pass-band ripple, sharp cut-off frequencies, and narrow transition bands was designed. Phase distortion—which can introduce time-delay—from each filter was eliminated by filtering the data, filtering the reverse of the filtered data, then reversing the output of the last operation of filtering.

Another consideration was that band-pass filtering would augment some broadband, physiological clatter that could be mistaken for a pathological fluctuation. It was subsequently presumed that an HFEO is approximately sinusoidal and that time-differentiation (whitening) of the IEEG would aid in discerning such waveforms. However, whitening would also enlarge ordinary oscillations within the IEEG. Therefore band-pass filtering and whitening were combined through a multiplicative operation as illustrated in Figure 5 to exploit the benefits of each pre-processing technique while reducing the respective disadvantages. Moreover, Figure 6 shows that this enhancement successfully emphasizes HFEO's in actual IEEG data.

Juxtaposing the time-aligned subfigures a-c of Figure 6 demonstrates the trade-offs that were discussed earlier about whitening and band-pass filtering. The red arrow in panel c was a high-frequency event highlighted through band-pass filtering that was not evident in the original (a) or time-differentiated (b) IEEG. Thus the product of the signals in plots b and c, which is shown

in panel d, diminishes the amplitude of the spurious event. On the other hand, the magenta arrows appear to be events that may have been too difficult to notice by the expert markers.

3.5 Processing: Feature Extraction and Genetic Programming

Feature extraction was the most important module of the method. Recall that feature extraction is the process of computing quantitative information (features) over a sliding window of data points, yielding a new time series of scalar quantities. Furthermore, effective classification of the data greatly depends upon the use of features that adequately discriminate events in an “HFEO class” and events in a “non-HFEO class.” The genetic programming algorithm was used to create a highly discriminatory feature from a selected set of features that were generated by classical feature extraction. Because the expected duration of the HFEO’s was between 100 and 800 milliseconds, a sliding window of 100 milliseconds (20 samples) was selected and shifted every 60 milliseconds (12 samples) to generate the matrix of classical features. The set of C-features, which is listed in Appendix B, was chosen based on low computational burden and ease in comprehension with at least one feature selected from conventional analysis domains (e.g., statistical, time). Arbitrary values of 35 generations and 350 individuals were assigned for the genetic programming procedure.

Figure 7 summarizes the general relationship between the genetic programming algorithm and the C-features, while Figure 8 demonstrates the powerful ability of the GP to improve class separation given the C-features. Reviewing Figure 6d, it could be assumed that an HFEO was characterized by activity with much more fluctuation and larger amplitude than moments without an HFEO. Thus it was expected that *curvelength* and *standard deviation*—two classical features that correlate with the amplitude and variation of a signal—would yield great separation between classes.

Figure 8a shows that some distinction between classes existed but was insufficient because several samples from the HFEO-class overlap with samples from the non-HFEOclass, especially in the non-HFEO region of the feature-space. The deficiency in separation by the two manually selected features raised the complicated question of what supplementary C-features were necessary for satisfactory separation. In contrast, Figure 8b exemplifies a preferred separation of classes in that most values for the features of each class are concentrated in two distant, disjoint intervals of the real number line— between .45 and .50 for the HFEO’s and zero for the non-HFEO’s, which was obtained automatically via the genetic program. Hence the GP algorithm produces a single feature with adequate class-separation in a more efficient manner than relying the physics of the problem and visual inspection.

The same routine was applied to each patient with the same parameters for the genetic programming algorithm. The GP outputted two features with appreciably different structures (trees) and fitness values. As shown in Figure 9, the tree of the GP-feature for the pediatric training sample (b) had more branches and nodes for functions (e.g., *max*, *times*, *cube*) than the tree of the GP-feature for the adult training sample (a). Correspondingly, the complexity of the trees translated to the complexity of the following equations defining the two features, where all operations (e.g. *division*, *square-root*) are protected as in Table A-1:

$$\begin{aligned}
 f_{child}(X4, X5, X7, X11) &= f_{child}(MeanPsd, Mean, Iqr, RenyiEntropy) \\
 &= \sqrt{\max(\log(\max(\log(\frac{d^2}{dt}(X4 \cdot X5^3))), \arctan(((X4 \cdot X5^2)^3 \cdot X5^3))), \arctan(\min(X7, |X4 \cdot X5^2| - \frac{X7}{X11})^9))} \\
 \\
 f_{adult}(X4, X5, X7, X9) &= f_{adult}(MeanPsd, Mean, Iqr, ShannonEntropy) \\
 &= \arctan(\arctan(\arctan(X4 \cdot \min(\arctan(X4 \cdot X5)^3 \cdot ((X7 - X5) \cdot X9)^3, \min(|X5|, \arctan(X9))))))
 \end{aligned}$$

Figure 9 (c–d) illustrates that distinguishing classes (i.e., HFEO and non-HFEO) should be easier for the adult data than the pediatric data according to the probability density function

(PDF) for each class per patient. This result was interesting because apparently a more complex tree did not produce better class-separation but was expected because the observation coincided with a visual review of training data, in which the contrast between an HFEO and non-HFEO was greater in the adult database than in the pediatric database. It was conjectured that the challenges in achieving equally high fitness values could be attributed to some physiological or pathological differences between the HFEO's of each patient.

3.6 Processing: Classification and Detection

A classifier that makes no assumptions about the form of the underlying probability density functions and the decision-boundary between classes of the given data while possessing relatively low computational complexity in both the training and testing phases was preferred for the detector. In view of that goal, a k-nearest-neighbor (KNN) rule was the most suitable classifier for the HFEO-detector. Applying the GP-feature that was created for each patient and the k-nearest-neighbor classifier to the testing data, the detector produced a binary sequence denoting the perceived presence (1) or absence (0) of an HFEO, or the classification of the IIEG that was inputted to the detector. Regarding the actual detection of an event, the beginning and ending times of an automatically marked signature were declared as the closest times at which the binary output transitioned from 0 to 1 and 1 to 0 respectively and stored for future evaluations.

3.7 Post-processing

The main goal of the post-processing stage was to remove events that were possibly bad classifications from a list of detections. The following two suppositions were made concerning the events that were automatically detected using the previously described modules:

1. Some events would be false-positives, or events that were not HFEO's, with the remaining events being true-positives; and
2. Although apparently similar to the desired events, false-positives should differ somehow in a feature-space from correct detections (true positives or true HFEO's).

Consequently, two questions were considered to improve classification accuracy by avoiding false-positives: 1) "What feature(s) could reveal a disparity among spurious and true events?" and 2) "How can a post-processing module exploit the knowledge of this disparity to prevent or reduce the number of false detections for the overall method?"

Choosing a small subset of records that were processed, resulting detections were labeled accordingly as either a TP or a FP and compiled. Although training a GP-feature to discriminate the true-positive (TP) and false-positive (FP) classes was considered as a solution to the first question, it was desired to investigate the GP-feature that was computed for the processing module. Figure 10 demonstrates that TP's and FP's overlap in the feature-space in a fashion similar to the overlap of the PDF's for HFEO's and non-HFEO's in Figure 9c. The above figure portrays the possibility of misclassifying a non-HFEO that may be similar to an HFEO according to the GP-feature. To remedy this issue—and answer the second question—it was decided to incorporate another level of classification based upon the above small samples of data. This post-processing stage computed a GP-feature value for each detected event and executed a *nearest-neighbor rule* for each value to identify potential false-positives, which subsequently were removed from the current list of detections. The main disadvantage of this approach is that true positives may be removed, thereby reducing the amount of available information for statistical analyses. However, if the TP's exhibit a high incidence rate or the benefit of eliminating FP's, which can taint the information needed for statistical analyses, outweighs the risk, then this drawback may be negligible.

4. Statistical Analyses

The performance of the completed detector as described in the earlier sections but without the post-processing stage was evaluated for each patient using a set of testing data containing $n = 20$ clippings of short-duration IIEG. After the detector was applied to the testing files, the automated markings were manually examined to register the times of any false positives and false negatives along with the times of the true positives that were automatically detected. The inventory of markings was then used to calculate the measures of performance (e.g., sensitivity, specificity) of the method for each record of the testing data, yielding samples of 20 values per measure from which confidence intervals in performance were computed. In addition, the proposed technique was compared with standard means to select and transform features for further evaluation. Thus, the above procedure was repeated for the following three benchmarks:

- The C-features selected by the *genetic programming* algorithm without the nonlinear transformation determined by the GP.
- Three C-features selected via an algorithm based on *forward sequential selection* (FSS). In FSS, an algorithm initializes an empty subset of features and evaluates a fitness (e.g. class-separation, error in classification) for all subsets of features with only one feature, choosing the subset with the best fitness. Repeatedly, a new feature that increases the fitness more than the other remaining features is added to the current subset until no significant improvement in fitness results.
- Two C-features selected by *visually inspecting* a subset of equations presumed to be quality considering the physics of the problem (and the literature).

Figure 14 illustrates the abovementioned techniques. All four methods relied on a nearest-neighbor rule as a classifier. The set of selected features for each benchmark were normalized for both the training and testing samples of the classifier in a “mean-variance” mode, or more specifically, a subtraction of the sample mean before division by the sample deviation, which was the most common approach to normalization of data in literature.

An analysis of variance (ANOVA) aptly tested if a statistically significant difference existed between the proposed method and the benchmarks, where the null hypothesis claimed that all methods for detecting epileptic oscillations performed equally in terms of sensitivity, specificity, positive predictive value, and negative predictive value; and the alternate hypothesis claimed that at least one method differed from the others per metric. The ANOVA for the pediatric patient determined no statistical equality (P -values $< .05$) between all of the methods for each of the four measures, while the ANOVA for the adult patient determined a statistical equality (P -value $> .05$) between all of the methods for only the measure of positive predictive value.

A *multiple comparison procedure*, specifically a *Tukey's least significant difference procedure*, provided information about which pairs of methods significantly differed and which were essentially equal in performance. Figure 11 represents the observed similarity between the approaches according to the Tukey's procedure, where the distance between a pair of nodes (benchmarks for feature-selection) is inversely proportional to the similarity. The results revealed that forward sequential search (FSS) and genetic programming (GP) performed similarly and that the detector using a GP feature outperformed the detector using features from a sequential search in the most cases across both patients. Additionally, figures 12 and 13 illustrate the relative confidence (at the level of 95%) in the performance of the methods for the adult and pediatric testing data, respectively. Detectors relying on both the FSS and the GP exhibited very high (greater than 96 percent) measures of specificity and negative predictive value. The GP feature demonstrated satisfactory values (greater than ~80%) for the median

sensitivity and PPV despite a large spread; while the FSS feature set performed equally regarding sensitivity but yielded a very low median PPV (70%) for the adult data.

In summary, the three previous figures show that feature-selection via GP was more robust than FSS and performed well. Moreover, it was recognized that choosing a GP over FSS promoted the advantages of reducing the dimension of the feature-space to a single dimension, which facilitates more efficient training and testing processes for a classifier, and controlling the class-separation in the feature-space.

5. Discussion & Conclusion

An automated approach to detect high-frequency epileptiform oscillations was successfully designed using genetic programming to create optimal artificial features. Overall the algorithm achieved superb measures of performance at the 95% level of confidence. Notwithstanding marginal values of 78.5% for the median positive predictive value of the adult testing data and 83.5% the sensitivity of the pediatric testing data, the binary detector boasted values between 93% and 99% for the other metrics across both patients. This work demonstrated that a binary detector including GP and a k-nearest-neighbor rule for classification could outperform similar detectors relying on other approaches for selecting features.

The primary advantage of using the GP to craft a single feature was that optimal class-separation was controllable for the problem, whereas using other techniques for feature-selection yields an arbitrary multi-dimensional set of C-features for which static, sub-optimal class-separation and higher computational burden for the classifier were tolerated. Moreover, with a single feature rather than a set of features—with a typically ambiguous constraint in size—the *curse of dimensionality*, which refers to the key fact in pattern-classification that the number of data samples required to estimate some arbitrary multivariate probability distribution increases exponentially as the number of dimensions in the data increases linearly, was directly addressed. The GP-feature computed for the pediatric training data was more complex than the feature computed for the pediatric adult. This finding unsurprisingly corresponded to the less obvious separation between an HFEO and background in the IEEG upon visual review of the data, but raised questions on the dynamics of epilepsy in adults and children that warrant further investigation.

In closing, this section presented work that advanced the state of the art in the detection and analysis of seizure precursors by developing a reliable general scheme for a precursor-detector. With this method better progress can be made toward reliably presenting statistics that aid in deciphering the mechanisms of epilepsy and establishing a methodical solution to isolating appreciably epileptic brain so that precise treatment and thorough understanding of epilepsy may be achieved. In future work, the proposed method will process seizure records with more IEEG channels and multiple patients to further study the spatial-temporal characteristics of HFEO's and investigate the existence of any reproducible outcomes that diagnose regions of epilepsy or predict the location or time of a seizure onset.

Acknowledgements

This work was partially funded by the Harriet G. Jenkins Pre-doctoral Fellowship Program, Whitaker Foundation, NIH, Dana Foundation, Epilepsy Foundation, and Klingenstein Foundation. The authors are grateful to the epileptologists at the University of Pennsylvania, the Children's Hospital of Philadelphia, and the Mayo Clinic in Rochester, MN for supplying annotated data.

References

- Adjouadi M, Sanchez D, Cabrerizo M, Ayala M, Jayakar P, Yaylali I, Barreto A. Interictal spike detection using the Walsh transform. *Biomedical Engineering, IEEE Transactions on* 2004;51:868–872.

- Clavagno, G.; Ermani, M.; Rinaldo, R.; Sartoretto, F. Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on.; 2000. 3586. A multiresolution approach to spike detection in EEG; p. 3582-3585.
- Firpi, H. Electrical and Computer Engineering. Michigan State University; East Lansing: 2005. On Prediction and Detection of Epileptic Seizures by Means of Genetic Programming Artificial Features.
- Firpi, H.; Goodman, E.; Echaz, J. Epileptic Seizure Detection by Means of Genetically Programmed Artificial Features. GECCO '05. ACM; Washington, DC: 2005a.
- Firpi, H.; Goodman, E.; Echaz, J. On Prediction of Epileptic Seizures by Computing Multiple Genetic Programming Artificial Features. In: al, MKe, editor. Eight European Conference on Genetic Programming, Lecture Notes in Computer Science. Springer-Verlag; Berlin Heidelberg: 2005b. p. 321-330.
- Hassanpour, H.; Mesbah, M. Neonatal EEG seizure detection using spike signatures in the time-frequency domain. Signal Processing and Its Applications, 2003; Proceedings. Seventh International Symposium on; 2003. p. 41-44.
- Hassanpour, H.; Mesbah, M.; Boashash, B. EEG spike detection using timefrequency signal analysis. Acoustics, Speech, and Signal Processing, 2004; Proceedings. (ICASSP '04). IEEE International Conference on; 2004. p. V-421-V-424.
- Jones, RD.; Dingle, AA.; Carroll, GJ.; Green, RD.; Black, M.; Donaldson, IM.; Parkin, PJ.; Bones, PJ.; Burgess, KL. A system for detecting epileptiform discharges in the EEG: real-time operation and clinical trial. Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine; Proceedings of the 18th Annual International Conference of the IEEE; 1996. p. 948-949.
- C-W, Ko; Lin, Y-D.; Chung, H-W.; Jan, G-J. An EEG spike detection algorithm using artificial neural network with multi-channel correlation. Engineering in Medicine and Biology Society, 1998; Proceedings of the 20th Annual International Conference of the IEEE; 1998. p. 2070-2073.
- Koza, JR. Hierarchical genetic algorithms operating on populations of computer programs; Proceedings of the 11th International Joint Conference on Artificial Intelligence; 1989. p. 768-774.
- Koza, JR. Genetic Programming: On Programming of Computers by Means of Natural Selection. MIT Press; Cambridge, MA: 1992.
- Koza, JR. Genetic Programming II: Automatic Discovery of Reusable Programs. MIT Press; Cambridge, MA: 1994.
- Marchesi, B.; Stelle, AL.; Lopes, HS. Detection of Epileptic Events using Genetic Programming. 19th International Conference-IEEE/EMBS; Chicago, IL. 1997. p. 1198-1201.
- Marsh, E. The Observed Bandwidth of High-frequency Epileptiform Oscillations in Pediatric EEG's. Smart, OL., editor. Philadelphia, PA.; 2005.
- Niederhauser JJ, Esteller R, Echaz J, Vachtsevanos G, Litt B. Detection of seizure precursors from depth-EEG using a sign periodogram transform. Biomedical Engineering, IEEE Transactions on 2003;50:449-458.
- Ossadtchi, A.; Leahy, RM.; Mosher, JC.; Lopez, N.; Sutherling, W. Automated interictal spike detection and source localization in MEG using ICA and spatial-temporal clustering. Biomedical Imaging, 2002; Proceedings. 2002 IEEE International Symposium on; 2002. p. 785-788.
- Pon, L-s; Sun, M.; Sciabassi, RJ. The bi-directional spike detection in EEG using mathematical morphology and wavelet transform. Signal Processing; 2002 6th International Conference on; 2002. p. 1512-1515.
- Smart, OL.; Worrell, GA.; Litt, B.; Vachtsevanos, GJ. Automatic Detection Of HFEO From Focal Intracranial EEG Using Fuzzy C-Means Clustering. Biomedical Engineering: New Challenges for the Future; Proceedings of the 2004 Annual Fall Meeting (BMES); Philadelphia, PA. 2004.
- Smart, OL.; Worrell, GA.; Litt, B.; Vachtsevanos, GJ. Automatic Detection of High Frequency Epileptiform Oscillations from the Intracranial EEG of Patients with Neocortical Epilepsy. 2005 Technical, Professional and Student Development Workshop (TPS); Boulder, CO. 2005.
- Staba RJ, Wilson CL, Bragin A, Fried I, Engel J Jr. Quantitative analysis of high-frequency oscillations (80-500 Hz) recorded in human epileptic hippocampus and entorhinal cortex. Journal of Neurophysiology 2002;88:1743-1752. [PubMed: 12364503]
- Tarassenko L, Khan YU, Holt MRG. Identification of inter-ictal spikes in the EEG using neural network analysis. Science, Measurement and Technology, IEE Proceedings- 1998;145:270-278.

- Van Hese, P.; Boon, P.; Vonck, K.; Lemahieu, I.; Van de Walle, R. A new method for detection and source analysis of EEG spikes. Engineering in Medicine and Biology Society, 2003; Proceedings of the 25th Annual International Conference of the IEEE; 2003. p. 2455-2458.
- Van Hoey, G.; Vanrumste, B.; Boon, P.; D'Have, M.; Van de Walle, R.; Lemahieu, I. Combined detection and source analysis of epileptic EEG spikes. Engineering in Medicine and Biology Society, 1998; Proceedings of the 20th Annual International Conference of the IEEE; 1998. p. 2159-2162.
- Wilson, K.; Webber, WRS.; Lesser, RP.; Fisher, RS.; Eberhart, RC.; Dobbins, RW. Detection of epileptiform spikes in the EEG using a patient-independent neural network. Computer-Based Medical Systems, 1991; Proceedings of the Fourth Annual IEEE Symposium; 1991. p. 264-271.
- Worrell GA, Parish L, Cranstoun SD, Jonas R, Baltuch G, Litt B. High Frequency Oscillations and Seizure Generation in Neocortical Epilepsy. Brain 2004;127:1-11.
- Zhang, T.; Yang, F.; Tang, Q. Wavelet based approach for detecting and classifying epileptiform waves in EEG. [Engineering in Medicine and Biology, 1999; 21st Annual Conf. and the 1999 Annual Fall Meeting of the Biomedical Engineering Soc.] BMES/EMBS Conference, 1999. Proceedings of the First Joint; 1999. p. 942

Appendix A: Proposed Use of the GP Algorithm

A genetic program can be applied for the optimal fusion of conventional features, or C-features. The input to the GP was an array of the C-features, the terminals are the subset of principal features chosen upon termination of the genetic program, and the functions are simple mathematical operations (e.g., addition, division, logarithm, maximum, cosine, square) that contribute to the nonlinear transformation of the terminals. The GP efficiently finds a multiple-input, single-output mapping, f , such that the output maximally contains information that distinguishes between classes of data.

$$\mathbf{y} = f(\mathbf{X}; \mathbf{c}) = f(\mathbf{X}^p)$$

The above expression describes the utility of a GP for feature-extraction, where \mathbf{X} is a matrix of feature-values $[\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_F]$ with each column-vector \mathbf{x}_i a distinct feature; \mathbf{X}^p is a matrix of feature-values $[\mathbf{x}_1^p, \mathbf{x}_2^p \dots \mathbf{x}_N^p]$ computed from the principal vectors selected by the GP, $\mathbf{x}_i^p = \mathbf{x}_j \forall i, 1 \leq i \leq F$, in \mathbf{X} ; \mathbf{c} is a column-vector of labels from the set $\{0, 1\}$ signifying the group of each sample of measures with each element corresponding to a row-vector in \mathbf{X} and \mathbf{X}^p and, f is the nonlinear transformation associated with the principal features, and \mathbf{y} is a column vector containing the values of the single, supreme feature.

With a training data set and predefined metric for the fitness of a prospective procedure for fusion, the GP iteratively tunes the parameters f and \mathbf{y} by imitating the four processes of natural evolution (i.e., *selection*, *crossover*, *mutation*, and *survival*) as described in Section 2. The following equations describe the fitness measure that is used in this work when the GP generates and sorts through various options for the optimal synthesis of features.

$$\begin{aligned} \text{fitness} &= (1 - \text{Pdf overlap}) \cdot (K \text{ factor}) \\ K \text{ factor} &= \frac{|\mu_1 - \mu_0|}{\sqrt{(\sigma_1^2 + \sigma_0^2)/2}} \\ \text{Pdf overlap} &= \int \min(p(\mathbf{y} | \mathbf{c} = 0), p(\mathbf{y} | \mathbf{c} = 1)) d\mathbf{y} \end{aligned}$$

In the equation for the *probability density function (PDF) overlap*, $p(\mathbf{y} | \mathbf{c} = c)$ is the PDF of the feature \mathbf{y} that is created by the genetic program, given that \mathbf{y} data belongs to a particular group c with label equal to 0 or 1. Values for the overlap range between zero and one. Although two groups with good separation, exhibit low overlap between the PDF's, low overlap alone does not guaranteed a feature with high discrimination. In the equation for the *K-factor*, the mean and variance for the PDF of each group is denoted by μ_c and σ_c respectively. Typically,

the higher the value for the *K-factor*, the more separated are the groups; however, this metric is not as accurate as the *PDF-overlap* given a feature with a multi-modal PDF for either group. Nonetheless, combining both metrics counteracts their individual disadvantages and produces an acceptable measure of fitness, referred to as the *overlap-k-factor product*, for blending a set of features. Furthermore, unlike fitness measures in previous work, the *overlap-k-factor product* is independent of the type of classifier and consequently provides more computational efficiency and generality in applying a GP for feature-fusion.

Furthermore, in genetic programming, the user must specify the function set and terminal set that the GP algorithm will use to create the programs. Appendix B discusses the conventional features that were candidates for terminals, and Table A-1 shows the functions. Since functions must satisfy the closure property in GP—not return values on which the GP cannot operate—certain functions (e.g., logarithm, division) are “protected,” or altered from their ordinary operation. The protected functions that were used in this work are acknowledged with their obvious adjustments in Table A-1.

Table A-1

List of the functions and their corresponding symbols and protection (if applicable) to maintain closure that was available to the GP algorithm of the proposed work.

Function	Symbol	Protection
Addition	+	N/A
Subtraction	−	N/A
Multiplication	×	N/A
Division	÷	Output 0 when denominator input is zero
Square	$()^2$	$()^2$
Cube	$()^3$	$()^3$
Square root	$\sqrt{ }$	Apply an absolute value operator before radical
Natural logarithm	Log	Output zero for an argument of zero; and Apply an absolute value operator to negative arguments
Absolute value		N/A
Sine	sin	N/A
Cosine	cos	N/A
Arctangent	atan	N/A
Maximum	max	N/A
Minimum	min	N/A

Appendix B: Library of C-Features

This section describes the list of conventional, or classical, features (C-features) that were candidates for the feature-selection component of the genetic programming procedure when the framework of GP was applied to feature-fusion for detecting an HFEO. Recall from the theory of GP that the selected candidates (features) are the terminals of the non-linear function corresponding to the GP-tree.

Let $n = \{v: v \geq (i-1)N + 1 \wedge v \leq iN, i \in \mathbb{N}\}$ be a set of time indices in samples and $f = \{v: v \geq 0 \wedge v \leq Fs/2\}$ be a set of frequency values in Hertz indices, where N is the number of elements in n , M is the number of elements in f , and Fs is the sampling rate for the EEG data. In addition, denote the IIEG sequence over a certain time interval i as $x(n)$, where $x = \{v: v_j = x(j) \forall j, 1 \leq j \leq N\}$, the Fast Fourier Transform (FFT) of $x(n)$ as $X(f)$, and the normalized FFT of $x(n)$ as $\hat{X}(f)$. Finally, let $x^{ord}(n)$ define the ordered values of $x(n)$, where $x^{ord} = \{x_1^{ord}, x_2^{ord}, \dots, x_N^{ord}\}$, and $pdf(x(n)) = pdf(x)$ where pdf is the probability density function (PDF) of a set of values. The following tables (B-1 through B-4) contain the preliminary library of conventional features that were considered as input to the genetic programming procedure for feature-fusion.

Table B-1

Time Domain Features

Feature	Equation	Description
Energy	$\frac{1}{N} \sum x(n)^2$	Average instantaneous energy
Nonlinear Energy	$\frac{1}{N} \sum (x(n)^2 - x(n-1)x(n+1))$	Change in amplitude and frequency
Curve Length	$\frac{1}{N} \sum x(n) - x(n-1) $	Length of the curve, similar to arc length or curvature
Crossings	$\frac{1}{N-1} \sum_{nn=n[2,3,\dots,iN]} ((\text{sign}(x(nn) - y^*) - \text{sign}(x(nn-1) - y^*)) > 0)$	Average number of intersections across an amplitude reference

Table B-2

Frequency Domain Features

Feature	Equation	Description
Maximum PSD	$\max_f X(f) ^2$	Maximum value of the estimated power spectral density
Mean PSD	$\frac{1}{M} \sum_f X(f) ^2$	Mean value of the estimated power spectral density
Peak Frequency	$f_{\max} \ni X(f_{\max}) = \max_f X(f) $	Frequency corresponding to the maximum value of the estimated power spectral density
Mean Frequency	$\frac{\sum_f f \cdot X(f) ^2}{\sum_f X(f) ^2}$	Frequency that is the centroid of power spectral density

Table B-3

Statistical Domain Features

Feature	Equation	Description
Mean	$\frac{1}{N} \sum x(n)$	Arithmetic mean or average of amplitude values
Standard Deviation	$\sqrt{\frac{1}{N-1} \sum (x(n) - \frac{1}{N} \sum x(n))^2}$	Unbiased standard deviation of amplitude values
Median	$\begin{cases} x_{\frac{N+1}{2}}^{ord}, & N \text{ is odd} \\ \frac{x_{\frac{N}{2}}^{ord} + x_{\frac{N}{2}+1}^{ord}}{2}, & N \text{ is even} \end{cases}$	Middle value when amplitude values are ordered from smallest to largest
Interquartile Range	$\begin{cases} \text{median}\{x_{\frac{N+1}{2}+1}^{ord}, \dots, x_N^{ord}\} - \text{median}\{x_1^{ord}, \dots, x_{\frac{N+1}{2}}^{ord}\}, & N \text{ is odd} \\ \text{median}\{x_{\frac{N}{2}+1}^{ord}, \dots, x_N^{ord}\} - \text{median}\{x_1^{ord}, \dots, x_{\frac{N}{2}}^{ord}\}, & N \text{ is even} \end{cases}$	Spread of the amplitude values

Table B-4

Information Theory Domain Features

Feature	Equation	Description
Shannon Entropy	$-\sum pdf(x) \cdot \log_2(pdf(x))$	Level of disorder, uncertainty, or randomness
Spectral Entropy	$-\sum \widehat{X}(f) \cdot \log_2(\widehat{X}(f))$	Level of disorder, uncertainty, or randomness
Reyi Entropy	$-\sum pdf(x)^p \cdot \log_2(pdf(x)^p)$	Level of disorder, uncertainty, or randomness
Complexity*	$\frac{\log_2(N)}{N} SequenceComplexity(x(n))$	Level of sequence complexity given by the number of unique subsequences within the sequence

*Note: The Complexity feature cannot be represented by a simple equation. In lieu of this difficulty, the algorithm *SequenceComplexity(x(n))* was written but the code is omitted. For more details, see Zhang et al, "Detecting Ventricular Tachycardia and Fibrillation by Complexity Measure" IEEE 1999.

Biographies

Otis Smart was born in Orangeburg, SC in 1978 and raised in Augusta, GA. He received the B.S. degree in general science from Morehouse College in 2001 and the B.S. and M.S. degrees in electrical and computer engineering from the Georgia Institute of Technology, Atlanta in 2001 and 2002, respectively. Presently, he is pursuing a Ph.D. degree in electrical and computer engineering at Georgia Tech. His thesis involves developing intelligent methods for the localization of epileptic tissue and the prediction and termination of epileptic seizures. His research interests include signal and image processing, data-mining, fuzzy clustering, and detection. Mr. Smart is NASA Harriet G. Jenkins Pre-doctoral Fellow.

Hiram Firpi earned his B.S. degree in electrical engineering from Polytechnic University of Puerto Rico in 1999, a M.S. in electrical engineering from University of Puerto Rico- Mayagüez in 2001, and a Ph.D. degree in electrical engineering from Michigan State University in 2005. He was a postdoctoral fellow with the Intelligent Control System Laboratory at Georgia Institute of Technology (2005-'06). He is now a postdoctoral fellow with the Indiana University-Perdue University at Indianapolis (IUPUI), where he is developing and implementing machine learning algorithms for bioinformatics and computational biology problems. His research interests range from pattern recognition and machine learning tools to biology-related problems and control system applications. He is a member of the Tau Beta Pi honor society.

George Vachtsevanos received the B.E.E. degree from the City College of New York, New York, NY, the M.E.E. degree from New York University, New York, NY, and the Ph.D. degree in electrical engineering from the City University of New York, New York, NY. He is currently a Professor in the School of Electrical and Computer Engineering at the Georgia Institute of Technology, Atlanta, where he directs the Intelligent Control Systems Laboratory. His research interests include intelligent systems, diagnostics and prognostics, and robotics and manufacturing systems. He has published in the areas of control systems, power systems, bioengineering, and diagnostics/prognostics. Dr. Vachtsevanos is a member of Eta Kappa Nu, Tau Beta Pi, and Sigma Xi.

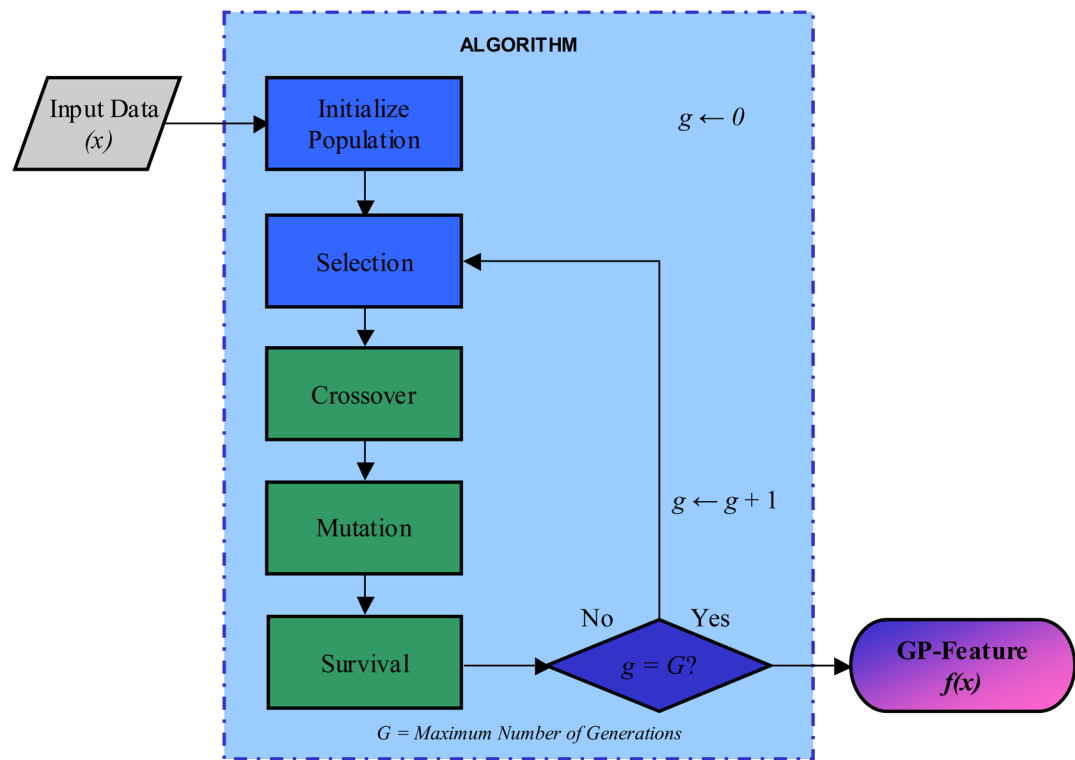
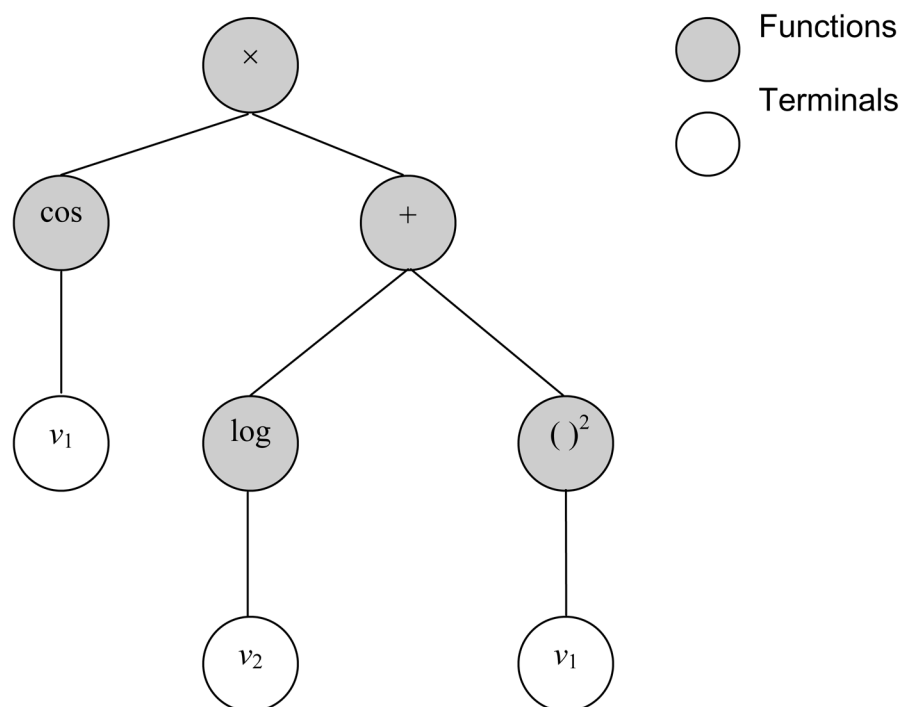
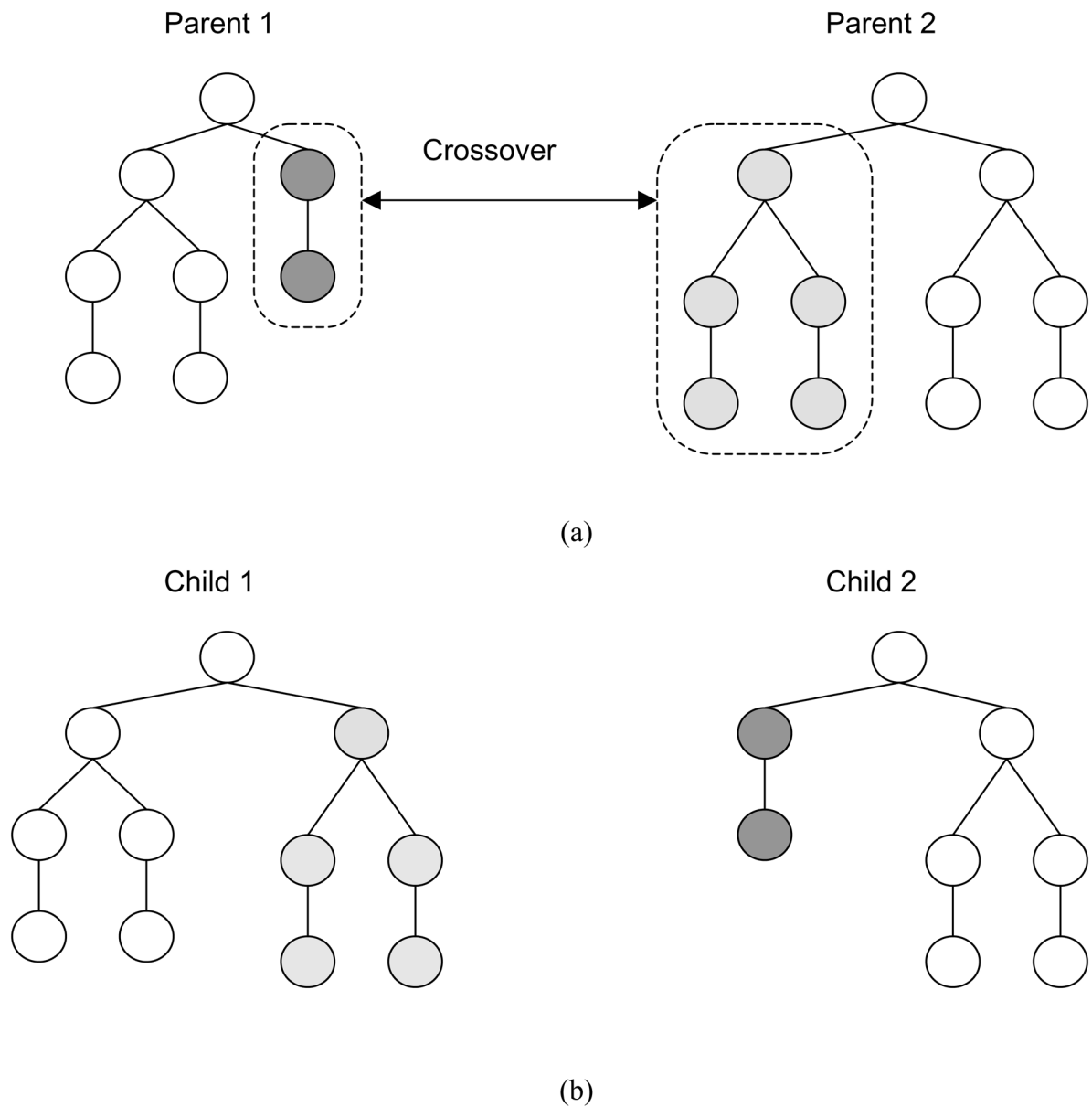


Figure 1.

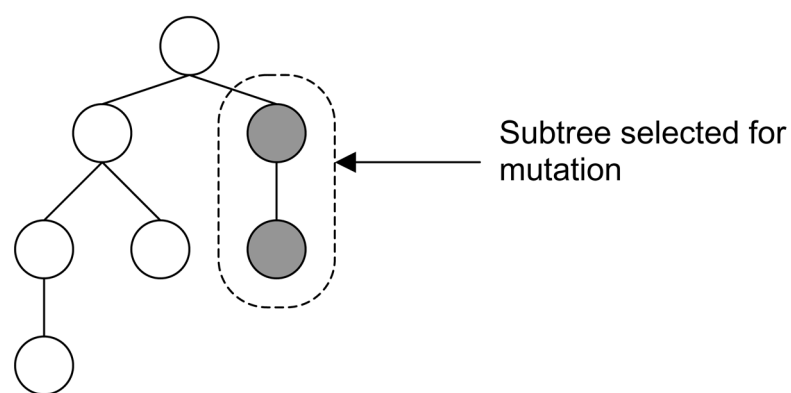
This figure illustrates the steps of a canonical genetic programming (GP) algorithm. The variable g denotes the number of generations, or iterations of the procedure, while the constant G is the desired number of generations. The result of the GP is a function—typically nonlinear—of the input data. The fitness measure depends upon the GP application.

**Figure 2.**

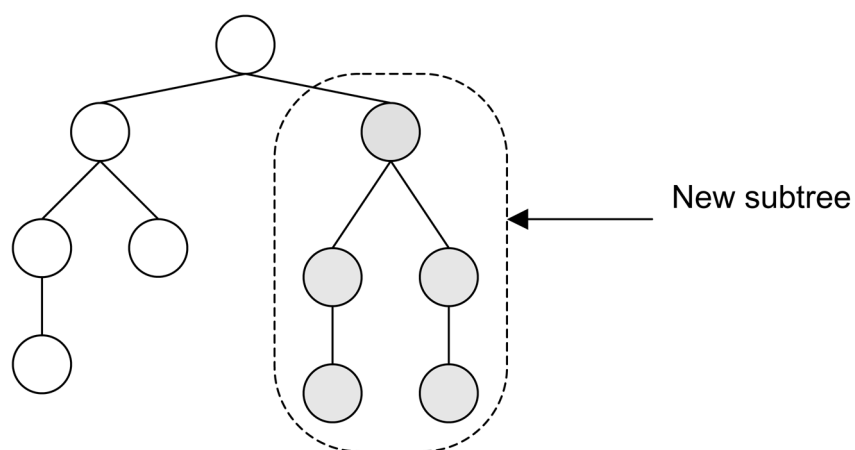
This figure exemplifies a solution from a particular GP procedure. The tree structure represents the following expression: $\cos(v_1) \cdot (\log(v_2) + (v_1^2))$.

**Figure 3.**

This figure demonstrates a typical cross crossover operator in a GP procedure: a) the over parents swap subtrees at random randomly selected nodes from each parent tree; and b) the children are the result after crossover is perform performed.



(a)



(b)

Figure 4.

This figure demonstrates a typical mutation operator in a GP procedure: a) a subtree is randomly selected from the node of a parent tree; and b) altering the subtree of the parent produces a child.

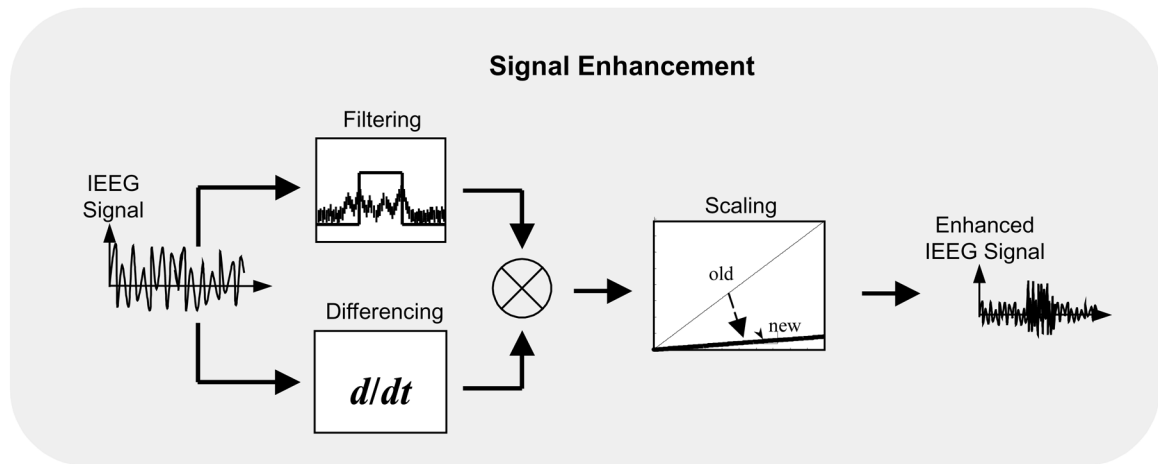


Figure 5.

This figure illustrates the pre-processing technique for highlighting an HFEO in a signal inundated with waveforms that are not high-frequency epileptiform oscillations. The enhancement is a simple multiplication of whitened and filtered data followed by a rescaling operation to preserve the true order of magnitude of the IEEG.

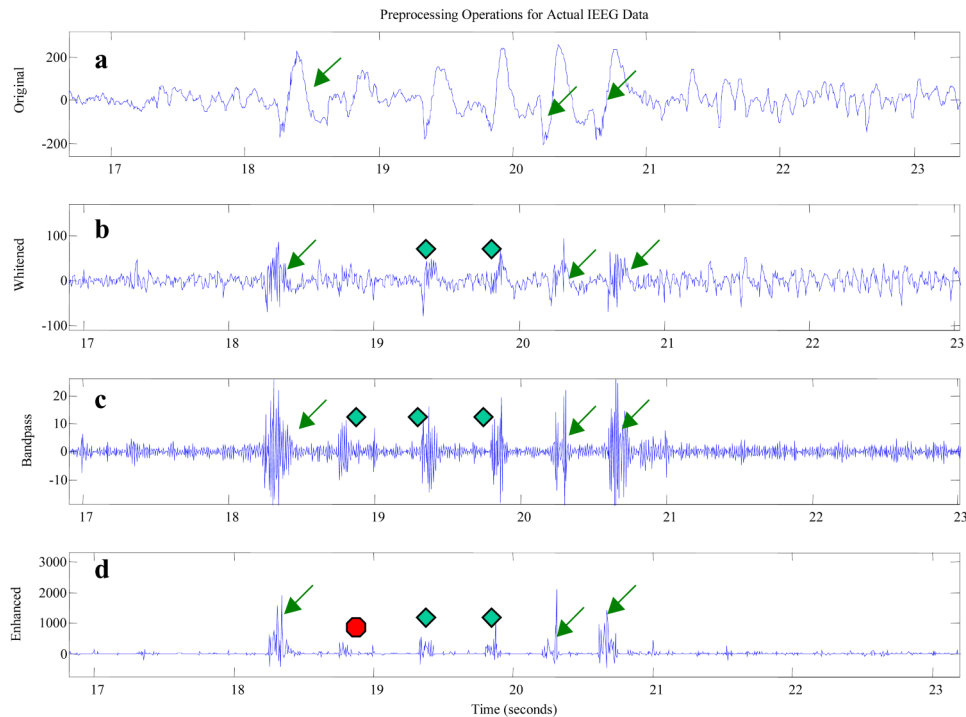
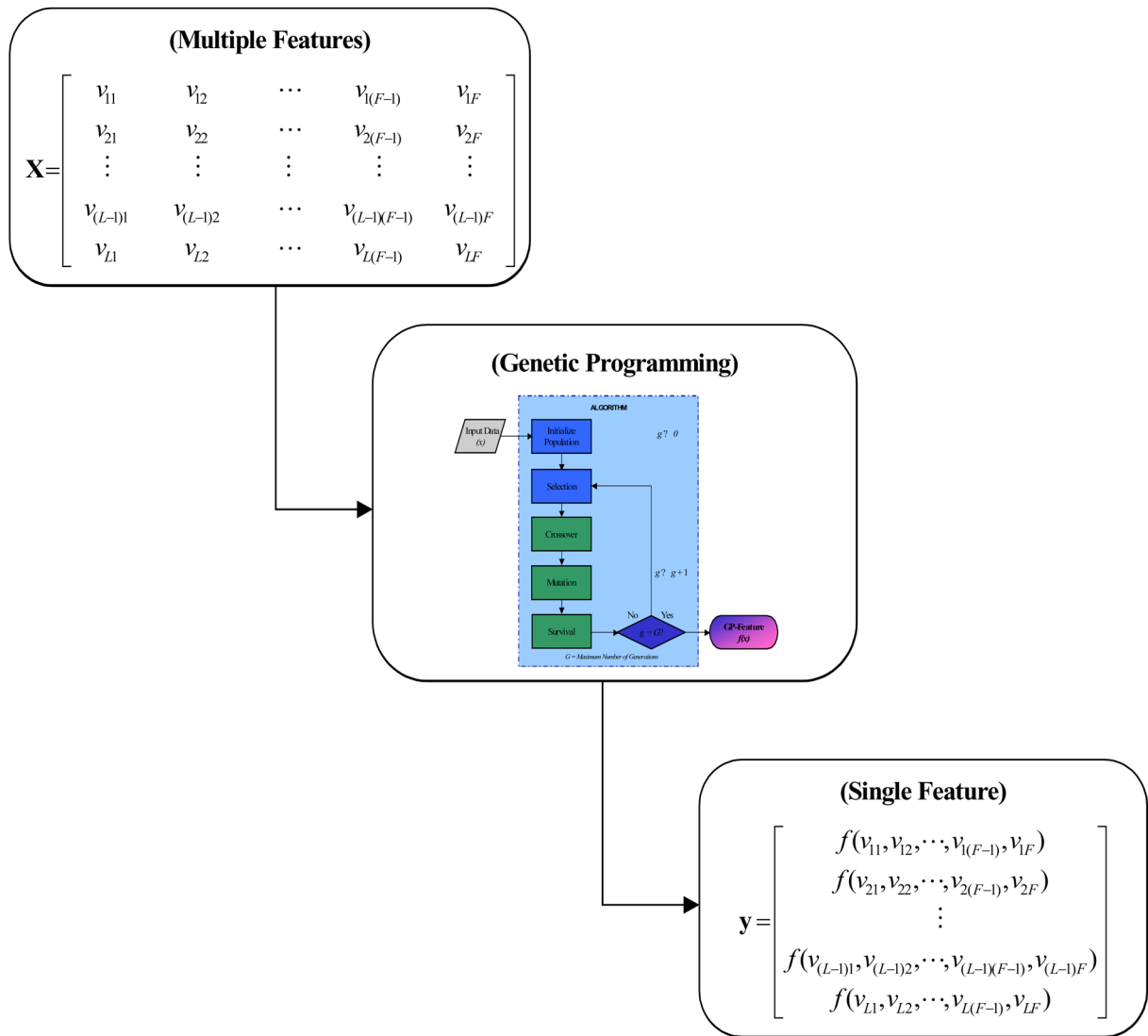
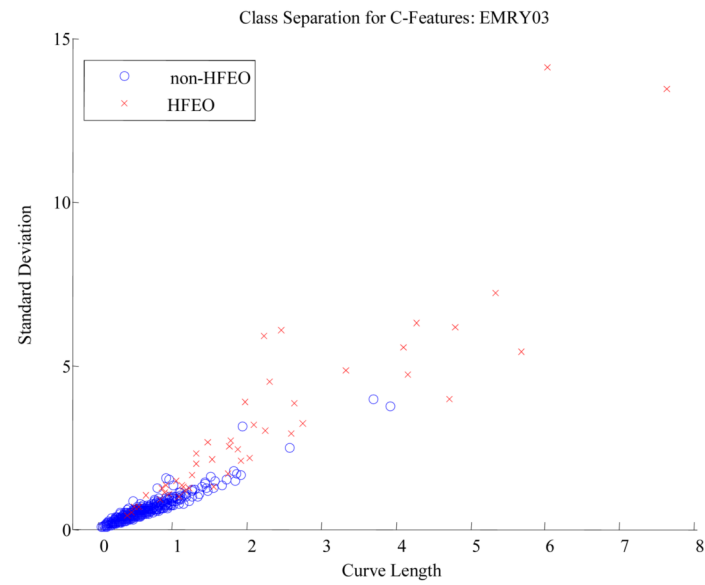


Figure 6.

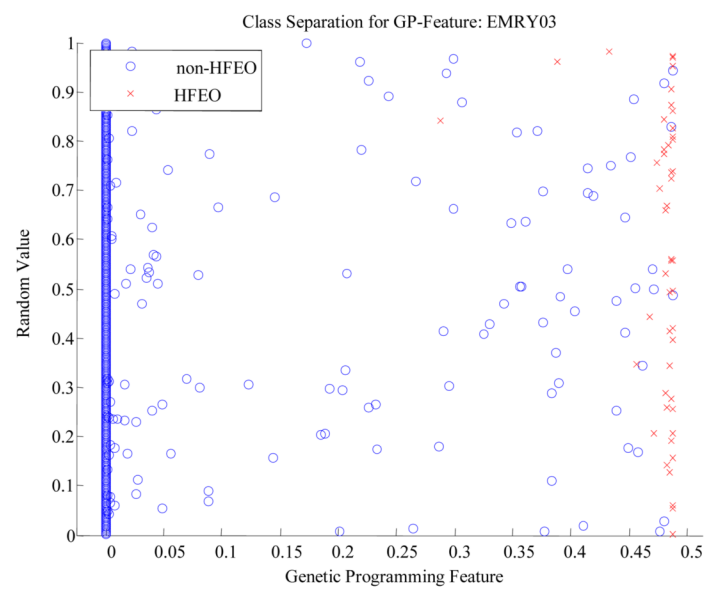
This figure shows the effectiveness of the pre-processing technique for emphasizing the activity of an HFEO contained within IEEG with low SNR (*a*). Although a trained reviewer identified three occurrences of an event (*arrows*), differencing (*b*), or whitening, and filtering (*c*) revealed two or three additional events (*diamonds*) in the unprocessed signal, respectively. The enhanced IEEG (*d*) contains six events that could be automatically detected: three known precursors (*arrows*), two potential false-negatives (*diamonds*) and one false-positive (*octagon*) due to filtering. Consequently, the suggested enhancement reveals the possibility that potential false positives (negatives) for the binary detector could actually be false negatives (positives) by the expert, giving value to the method still as a means to register putative events that may not be visually possible by the expert despite the few spurious events.

**Figure 7.**

This figure shows that the GP algorithm fuses a matrix (\mathbf{X}) of values computed from a set of distinct, multiple features into a one-dimensional vector (\mathbf{y}) for a single feature, f , which generally speaking is a nonlinear transformation of \mathbf{X} .



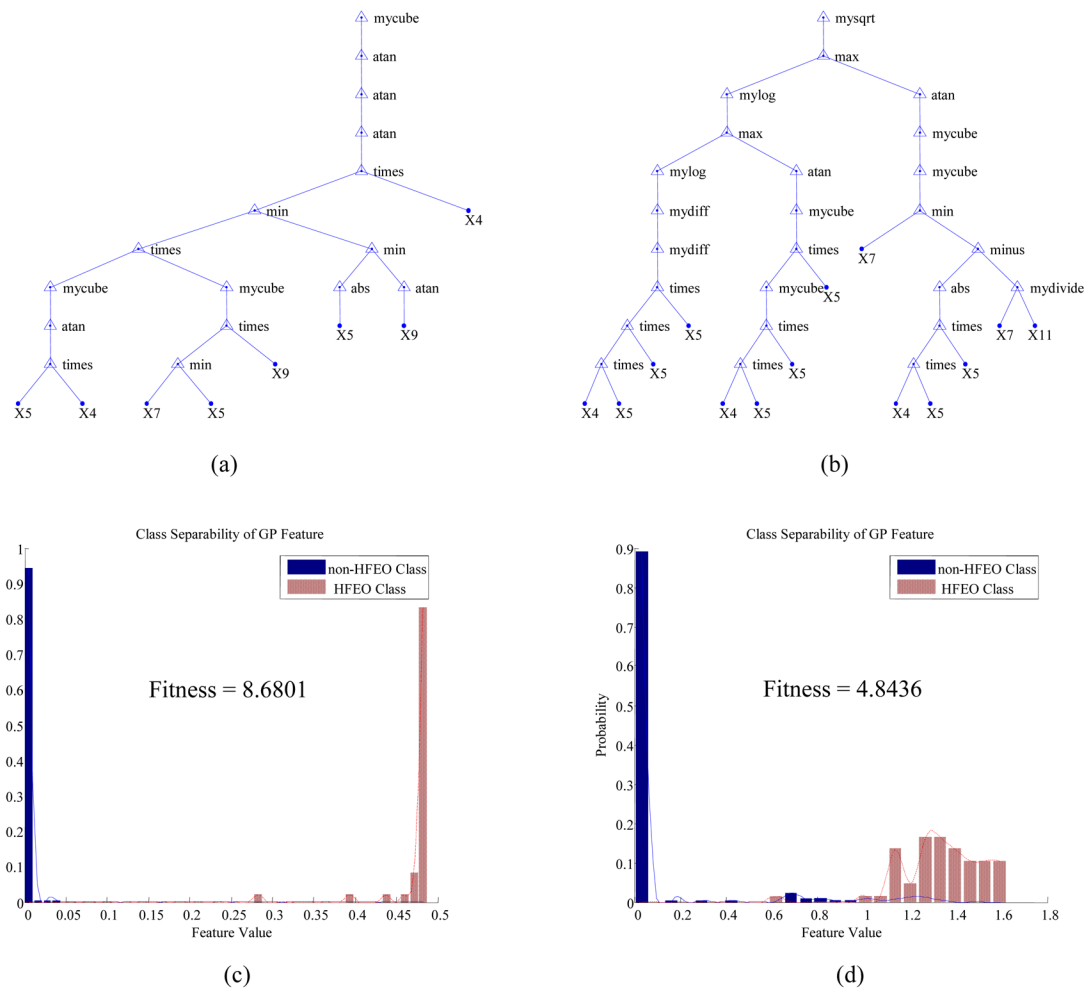
(a)



(b)

Figure 8.

This figure exhibits that a GP-feature (b) was more apt for the distinction between an HFEO and a non-HFEO than two intelligently selected features (a).

**Figure 9.**

This figure shows the trees and probability mass functions with corresponding fitness values obtained using the GP algorithm on training data for the adult (*a*, *c*) and the child (*b*, *d*) patients. A tree represents the nonlinear transformation of the C-features that were selected by the GP and the corresponding fitness value measures the level of discernment between binary classes based on the probability distribution of each class. Despite using the same parameters in the GP, a feature with a higher fitness and less complex tree resulted for the training sample of the adult patient. This was attributed to the very low signal-to-noise ratio (or contrast between background and oscillations) of the pediatric data.

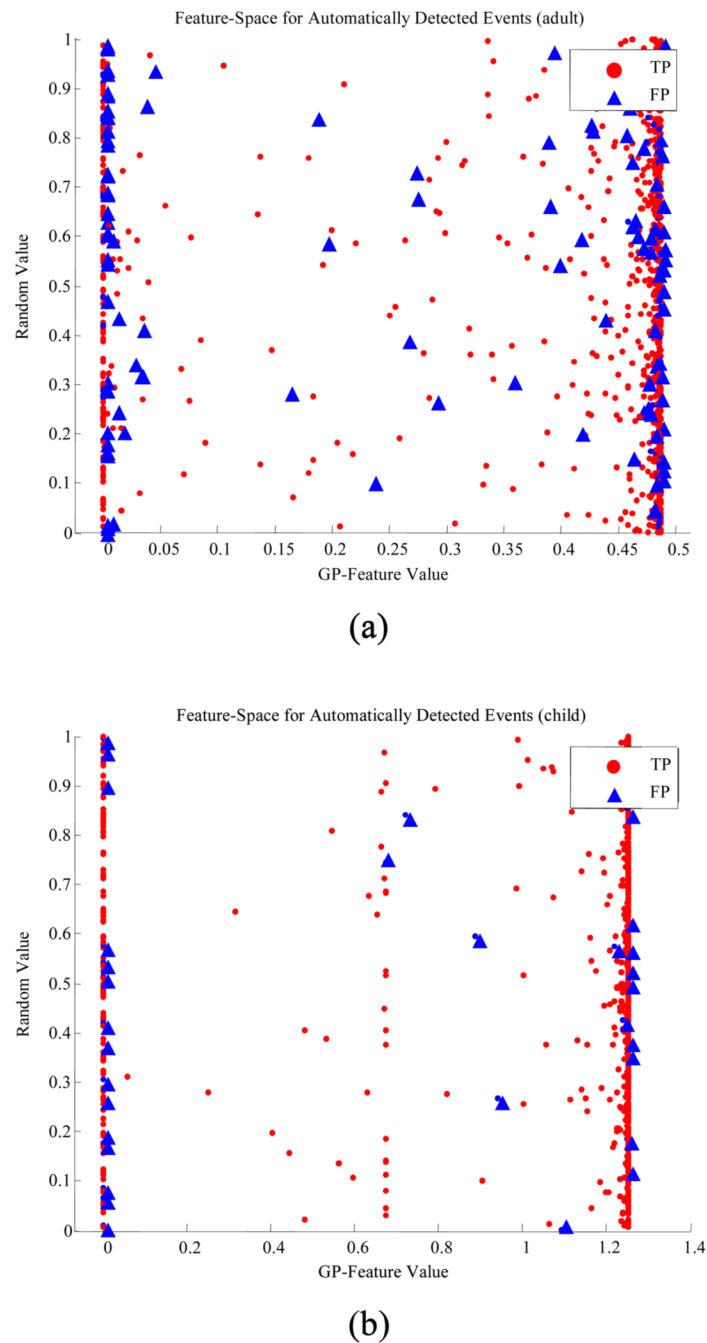


Figure 10.

This figure captures the feature-space for a sample of detected events from the adult (a) and pediatric (b) testing data. Each sample was stratified as a spurious (*triangle*) or a correct (*dot*) detection. The overlapping of the probability mass functions in Figure 9 leads to the overlapping of spurious and correct events above illustrated.

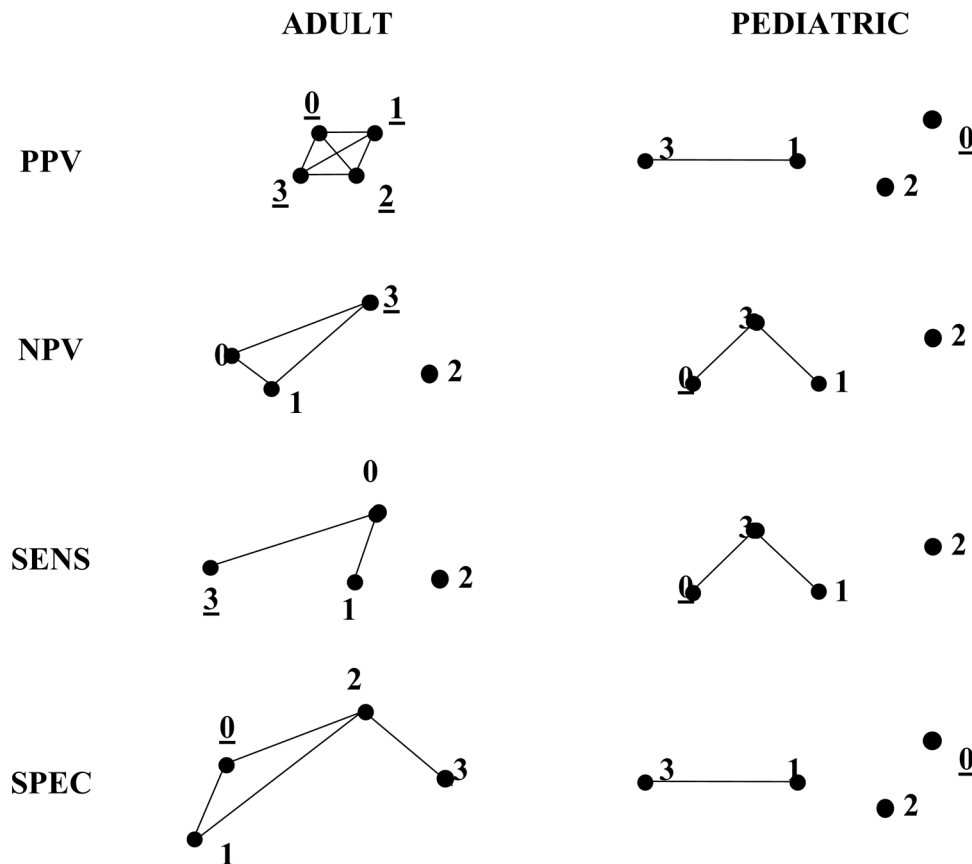


Figure 11.

This figure demonstrates the likeness between the methods for feature-selection (*lines between nodes*) as well as the method with the statistically highest measure of performance (*node with an underscore*). The figure is sketched to scale such that the distance between connected points is inversely related to the similarity between methods according to a Tukey's ranking test. Isolated nodes possess absolutely no similarity to any other method. The methods are symbolized by the following numbering: 0) features selected, transformed, and fused via GP; 1) features selected via GP and normalized in a mean-variance fashion; 2) features selected upon visual review considering the physics of the problem an assumption on the epileptic oscillations; 3) features selected via a forward sequential search algorithm.

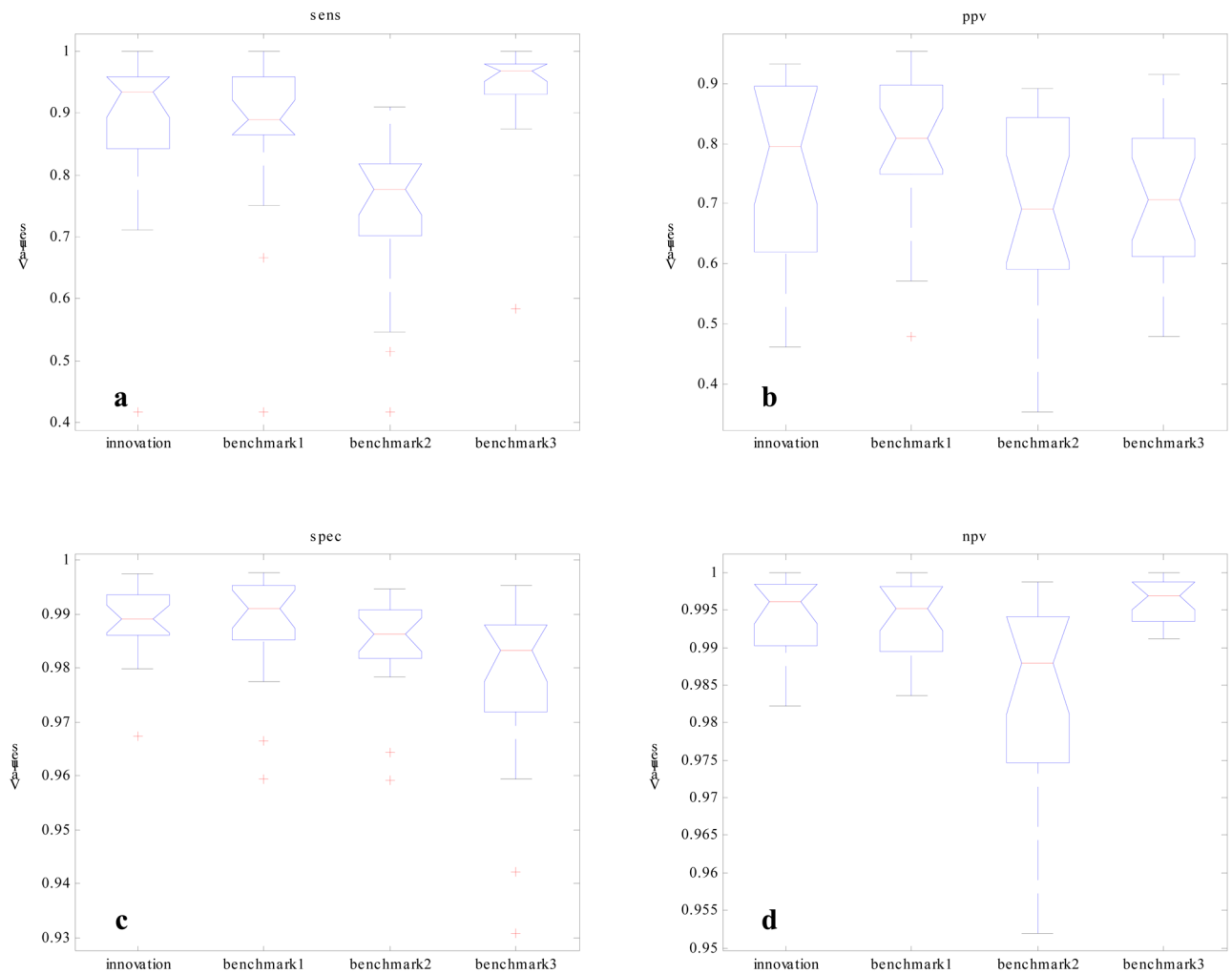


Figure 12.

This figure illustrates the performance of the HFEO-detector when verified using three types of feature-selection approaches for the adult testing data. For each subfigure a–d, the x-axis is the approach (from left to right: GP selection and transformation, GP selection and mean-variance normalization, “intelligent” selection, forward sequential search) and the y-axis is the value of a particular performance metric. Four metrics were considered: *sensitivity* (a), *positive predictive value* (b), *specificity* (c), and *negative predictive value* (d).

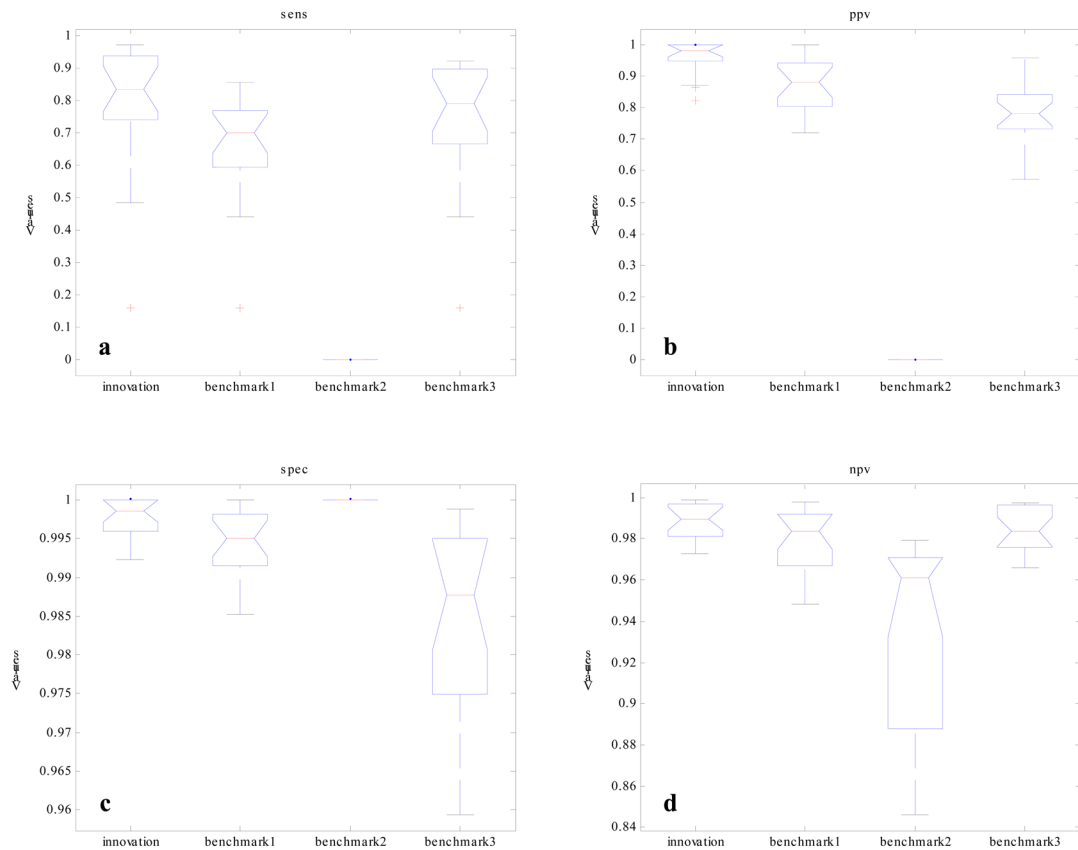
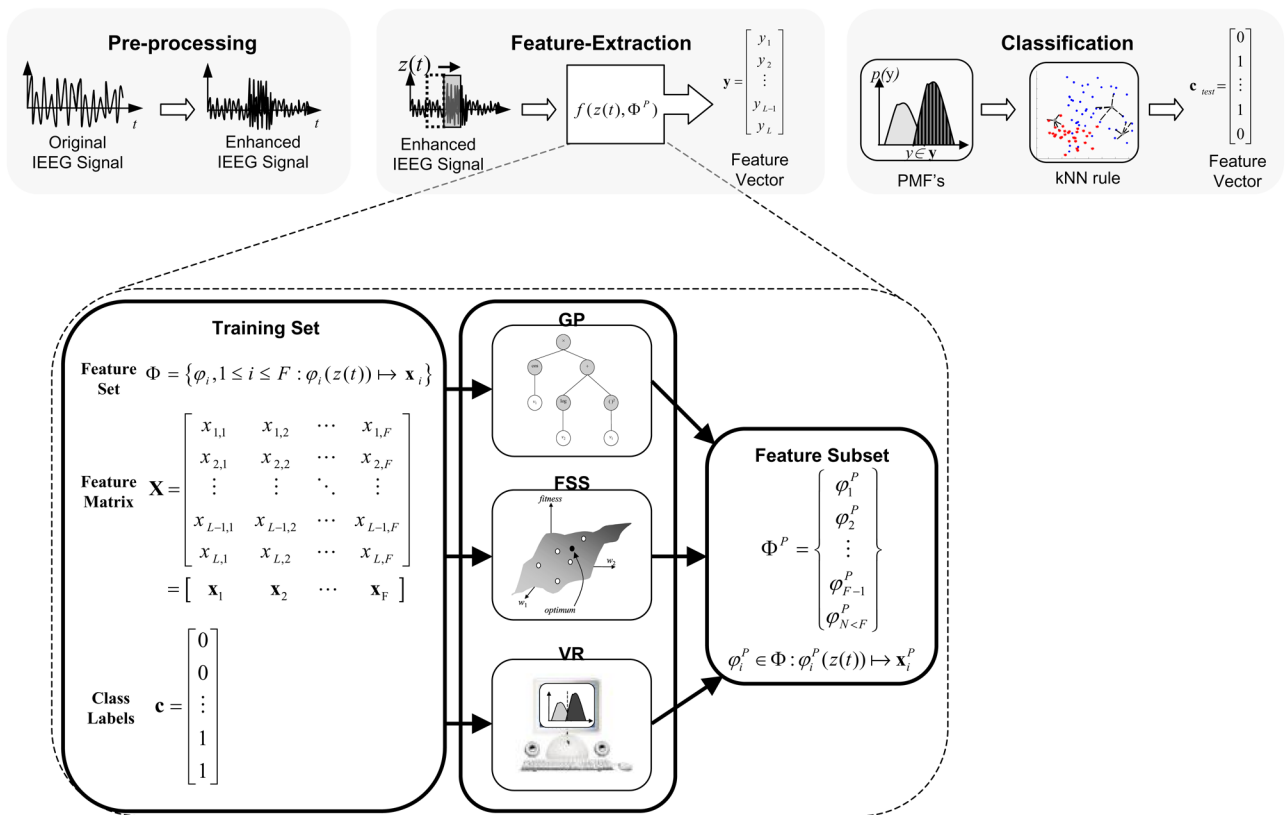


Figure 13.

This figure illustrates the performance of the HFEO-detector when verified using three types of feature-selection approaches for the pediatric testing data. For each subfigure a–d, the x-axis is the approach (from left to right: GP selection and transformation, GP selection and mean-variance normalization, “intelligent” selection, forward sequential search) and the y-axis is the value of a particular performance metric. Four metrics were considered: *sensitivity* (a), *positive predictive value* (b), *specificity* (c), and *negative predictive value* (d).

**Figure 14.**

This figure illustrates the role of three techniques to use features in the feature-extraction module of a binary detector for epileptic seizure-precursors: genetic programming (GP), forward sequential selection (FSS), and visual review (VR). According to a predefined measure of fitness, each technique returns the best subset of features, Φ^p , from a list of features, Φ , in a training set with a vector of values, \mathbf{x} , for each feature, \mathbf{x}_i , and a corresponding vector of labels, \mathbf{c} . Preferably, a rule or operation, f , to fuse a feature-matrix obtained after applying the feature in to an enhanced time-series of IEEG, $z(t)$, yields the vector, \mathbf{y} , to be classified such that a precursor (*one*) are discerned from background activity (*zero*).